DOCUMENT RESUME

ED 395 940 TM 025 007

AUTHOR	Pennock-Roman, Maria
TITLE	The Status of Research on the Scholastic Aptitude Test (SAT) and Hispanic Students in Postsecondary Education.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-88-36
PUB DATE	Jun 88
NOTE	62p.; Revision of a paper presented at a meeting organized by the Intercultural Development Research Association for the National Commission on Testing and Public Policy (San Antonio, TX, February 26, 1988).
PUB TYPE	Information Analyses (070) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Access to Education; College Bound Students; College Entrance Examinations; *College Preparation; College Students; Counseling; Course Selection (Students); Difficulty Level; Educational Research; Ethnic Groups; Grades (Scholastic); Higher Education; High School Students; *Hispanic Americans; *Item Bias; Language Proficiency; Parent Background; Performance Factors; *Predictive Validity; *Test Items; Test Results; Whites
IDENTIFIERS	*Hispanic American Students; *Scholastic Aptitude Test

ABSTRACT

This review of research studies focused on differences between Hispanic-American and White, non-Hispanic groups on the Scholastic Aptitude Test (SAT). Questions studied were the factors associated with ethnic group mean differences on the SAT; the types of item format or content found differentially easier or more difficult for Hispanics; the predictive validity of the SAT; and the adequacy of Hispanic students' test preparation. Data for the studies were from the 1987 College Board Profiles of College Bound Students and other College Board information. Mean differences between Hispanic and non-Hispanic White students were relatively large and were associated with differences in language background, parental education, high school grades, and type of academic courses taken. The numbers of items showing differential difficulty levels were small and not linked with differences in predictive validity. Overall, tests were slightly less accurate in predicting Hispanic students' success in college than for non-Hispanic Whites. The largest barrier to college access for Hispanic students may be inequity in the availability of guidance counseling. Although evidence concerning the adequacy of college admissions tests for Hispanic students is correlational, data do suggest room for improvement in both tests and test preparation. (Contains 4 tables and 46 references.) (SLD)

RR-88-36

ED 395 940

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

THE STATUS OF RESEARCH ON THE SCHOLASTIC APTITUDE TEST (SAT) AND HISPANIC STUDENTS IN POSTSECONDARY EDUCATION

Maria Pennock-Roman

Educational Testing Service
Princeton, New Jersey
June 1988

BEST COPY AVAILABLE

TM025007

The Status of Research on the Scholastic Aptitude Test (SAT) and Hispanic Students in Postsecondary Education[1,2]

Maria Pennock-Roman

Educational Testing Service, Princeton, NJ 08541

May 25, 1988

[1] Revision of a paper presented on February 26, 1988 in San Antonio, Texas at a meeting organized by the Intercultural Development Research Association for the National Commission on Testing and Public Policy. It was originally entitled The Status of Research on Selective Admissions Tests and Hispanic Students in Postsecondary Education. The paper and meeting were sponsored by the Ford Foundation and the University of California at Berkeley. This article will be published in a forthcoming volume of papers invited by the Commission.

[2] The author gratefully acknowledges Leonard Ramist's suggestions for the interpretation of the relationships between academic course taking and group mean differences in the 1987 Profiles data and many other comments. Henry Braun, James Deneen, Alicia Schmitt, Donald Powers, and Carol Slaughter also contributed valuable points in their reviews of this manuscript. Victor Bunderson and Hugh Cline provided the funds for the preparation of this manuscript.

Copyright © 1988. Educational Testing Service. All rights reserved.

Abstract

In this review of research studies on differences between Hispanic-American and White, non-Hispanic groups on the Scholastic Aptitude Test (SAT), four questions are addressed: (1) What factors are associated with ethnic group mean differences on the SAT? (2) What types of item format or content are identified as being differentially easier or more difficult for Hispanic vs. White, non-Hispanic students? (3) How accurately do selective admissions tests predict the performance of Hispanic students in college? and (4) Do Hispanic students have equal access to information necessary for long-term and short-term preparation for selective admissions tests?

The Status of Research on Selective Admissions Tests and Hispanic Students in Postsecondary Education

In this paper, studies evaluating the validity of the Scholastic Aptitude Test (SAT) for use in college admissions of Hispanic students will be reviewed. Other tests for admissions to graduate or professional schools are not considered here because small sample sizes and other practical problems limit the number of studies on Hispanic vs. non-Hispanic White group differences in validity. Before beginning, it is worthwhile to repeat the cautions in Linn's (1982) preface to his review of group differences in test validity.

> The controversies over testing are neither created by, nor will they be resolved by, the results of investigations of test validity (Cronbach, 1975) ... Justification of test use obviously depends upon much more than [how well an ability test predicts academic or professional performance]. Potential benefits and losses for the individual, the institution, and the society at large need to be considered, and the relative importance of the benefits and losses can be expected to vary greatly in the eyes of these various interests. Nonetheless, information about the degree of relationship of test scores to particular criterion measures and about the degree to which the observed relationship is generalizable across situations and from one situation to another is an important component in the evaluation of the use of tests ... (pp. 335-336).

In this presentation, I have taken a broad view of test validity, going beyond considerations of how well tests predict undergraduate grades. Group differences are reviewed in terms of the relationship between test scores and other educational and demographic variables, evaluations of test-item content and format, and the availability of college-admissions counseling and guidance information. Discussion of the studies will be organized around the following questions.

(1) Mean Differences. What factors are associated with ethnic group mean differences on the SAT? What are the implications of these differences?

(2) Evaluations of test-item content. Do test items contain material that is differentially more difficult for Hispanic students for reasons that are not relevant to the purpose of the test?

(3) Predictive validity. How accurately do selective admissions tests describe the performance of Hispanic students in college? When added to high school grades or other measures of achievement, do tests improve the identification of talented Hispanic students?

(4) Test Preparation. Do Hispanic students have equal access to information necessary for long-term and short-term preparation for selective admissions tests?

In evaluating the use of tests for selective admissions to higher education, it is necessary to consider all of these issues because mean differences alone are not sufficient to establish that tests are biased or that they represent unfair barriers to higher education. We have to determine carefully to what extent tests are giving us accurate information that is relevant to the decisions we have to make. While it is possible that the lower test scores could be due to content or item formats that lead to unintended cultural biases, it is also possible that the lower test scores may reflect real deficits in the quality of preparation Hispanic students have had for college. If so, then tests are serving the role of a messenger that merely conveys bad news, and killing the messenger will not solve the underlying problem.

One way to approach the issue of test bias is to consider the relative difficulty of individual items for different groups. This procedure (which is called differential item functioning or DIF) can isolate what items are potentially problematic because they contribute the most to group differences that are independent of overall ability level. While these methods are very

sensitive to group differences on individual items, for reasons to be explained later, they cannot be used to evaluate how much the test as a whole may or may not be biased.

Hence, it is necessary to examine also how well total test scores predict performance in college, which is the most direct way of evaluating the accuracy of information that the total test score provides. However, as will be explained later, there are many practical problems with validity research that limit its sensitivity to the detection and interpretation of group differences in the accuracy of measurement.

If we find evidence of differential validity, we need to establish why. It could be related to test content, which leads us again to differential item functioning (DIF) to determine the sources of these group differences in the items. On the other hand, the problem may be more pervasive, and not identifiable with a few isolated items. Alternatively, the source may be a lack of familiarity with standardized tests, independently of specific content. For this reason we need to examine the resources that Hispanic students have for preparation for tests, both long term and short term. However, as with mean differences, to demonstrate greater test naivete among Hispanic students is not enough to show bias because test naivete may also affect performance in college where grades are also partly based on test-taking skills. This takes us back to the issue of predictive validity and how accurately tests reflect future college performance. Hence all of these approaches are pieces of a puzzle, complementary in the picture they form on group differences in test performance.

In addition to summarizing existing studies and work in progress, desirable directions for new research will be suggested. As you will see, there is simply not enough information at the present time to make definitive

conclusions. Much research is in progress still, or remains to be initiated. The purpose of this paper is to outline what we do know and to suggest what questions we should be asking.

Mean Differences between Hispanic and White, Non-Hispanic (NH) Students and Variables Associated with Higher Test Scores

Data for this section are taken from the College Board 1987 Profiles of College Bound Students and a College Board press release entitled "National Scores on SAT Show Little Change in 1987; New Data on Student Academic Backgrounds Available" (9/22/87). Although these data are representative of college-bound students in states where most institutions require the SAT, they have several limitations. One limitation is that students from central, mountain, some southern, and some western states are not well represented in this data set because the American College Test (ACT) is more often the required college admission test for institutions in these regions. Also test results in these data are not generalizable to the overall high school student population because of self-selection to take the test and apply to college. The percentage of high school students who take the test varies by state, and the means for states with the largest proportions of examinees tend to be lower because a wider of variety of students attempt the test, and not just the very best students. While the data set contains extensive information on students' course taking patterns, it contains no ready measure of the quality of the courses and the high schools attended by the students, other than academic vs. non-academic categories.

For the study of Hispanic students' test performance, it is fortunate that the majority of states with large Hispanic concentrations (California, Texas, Florida, New York, New Jersey, and Pennsylvania) are primarily SAT

states, although several states with moderately large Hispanic concentrations (New Mexico, Arizona, Colorado, and Illinois) are primarily ACT states, and are therefore not well represented in the SAT files. However, these data have added limitations for studying college-bound Hispanic students. The profile of Puerto Rican students is less informative than it could be since it reports as one category residents from both the island (Commonwealth) of Puerto Rico and continental United States, groups that are quite distinct in language background and language of instruction. From 1981 to 1985, data reported on Puerto Rican examinees for the College Bound Profiles included only residents of continental U. S., and excluded Puerto Ricans residing in the Commonwealth. Regrettably, this distinction was not made in the data analyses for 1987 (personal communication, L. Ramist, January 1988), although the information on residence is available for 1987 and it was possible to have run the analyses separately for the two groups of Puerto Ricans. Unlike Puerto Rican students residing in continental U. S., Puerto Rican island residents (hereafter called Commonwealth) have usually learned English as a second language and have received much of their schooling in Spanish. The pattern of means for 1985 shows that Commonwealth Puerto Ricans have a lower SAT-V mean (352) than do continental Puerto Ricans (373) although the reverse is true for SAT-M --422 vs. 405, for Commonwealth and continental Puerto Ricans, respectively (personal communication, L. Ramist, 5/9/88). Thus it appears that Commonwealth Puerto Ricans have better developed skills in mathematics than do continental Puerto Ricans. Although the lower English proficiency of Commonwealth Puerto Ricans depresses their scores in both subtests, this effect is more evident in the the SAT-V.

Another limitation of the SAT data base is that the classification of race/ethnicity is incomplete because 5% of students did not fill out the

optional Student Descriptive Questionnaire which contains the item on self-classification by race/ethnicity in 1987, and an additional 1.7% left the race/ethnicity question blank. While this response rate is much higher than that usually found in social science surveys, studies on race/ethnic groups should be interpreted with caution because 6.7% missing values on this question is still a large number of individuals --about 72,000. Nevertheless, these data provide a comprehensive view of college bound students in the majority of states in the union.

There are several questions of interest, and each in turn is presented below.

1. How do overall mean SAT scores for Hispanic groups compare with those of the White, non-Hispanic (NH) group in 1987?

Insert Table 1 about here.

The means for SAT Verbal and Mathematical scores are shown in Table 1 by year and racial/ethnic group, from 1976 through 1987. Results are reported separately for Mexican-American, Puerto Rican, and Other Hispanic (also referred to as Latin American) groups. However, separate means for this latter group are available only for 1987. In other years, students in this category were included in the "White" or "Other" classifications. For 1987, the latest year, we find that in comparison to White NH students the Verbal (V) and Mathematics (M) means are substantially lower. The largest differences are found for the Puerto Rican group, 87 (V) and 89 (M) points, and the smallest differences are found for the Latin-American group which scored 60 (V) and 57 (M) points lower than White NH students. The Mexican-American group scored 68 (V) and 65 (M) points lower than White NH students.

The mean differences between Black and White NH students are larger (96 (V) and 112 (M) points) than the above differences for the Hispanic students.

2. Are there any changes in these differences over time since 1976?

There are some noticeable changes in means over time, but they must be interpreted with caution, because the data are cross-sectional and not longitudinal in nature. That is, changes may represent differences in self-selection trends (i.e., who decides to take the test within each group), and not necessarily improvements or decreases in mean skill levels of groups. For White NH students, Verbal scores decreased 9 points between 1976 and 1980, then slowly climbed up 7 points between 1982 to 1985, but currently are still slightly lower than in 1976. A similar pattern is found for Mathematics scores for White NH students. For Black students, SAT-V and SAT-M scores show very little change from 1976-1979, and then there is a steady increase that begins in 1979 for SAT-M and in 1981 for SAT-V that continues through 1987. For Black students the 1987 Verbal and Mathematics means are 19 and 23 points higher, respectively, than in 1976. A similar pattern is found for the Mexican-American group, except that the gains are smaller, there is a marked drop in SAT-M scores in 1978, and the 1987 SAT-V and SAT-M means are lower than in 1985. The Verbal (V) and Mathematics (M) means in 1987 are 8 (V) and 14 (M) points higher than in 1976 and 9 (V) and 22 (M) points higher than in 1978. The Puerto Rican group shows a pattern that is more like the White NH trend because there is a steady decline in both SAT-V and SAT-M scores from 1976 to 1978 or 1979, with an upturn beginning in 1980. The highest means are found in 1985 which are 23 (V) and and 21 (M) points than the lowest means in 1979. Hence, the decreases and increases are steeper than those for the White NH group and the 1987 means are lower than in 1985. As a result, there is

BEST COPY AVAILABLE

very little net change between the 1987 and 1976 means. The 1987 SAT-V is only 4 points lower than in 1976, whereas the SAT-M is 1 point higher.

Overall, differences in SAT-V and SAT-M means between White NH students and Black and Mexican-American students have narrowed in the last 11 years. For Black students the differences in 1976 were 119 and 139 points; in 1987 they were 96 and 112. For Mexican American students the differences in 1976 were 80 and 83 points respectively, and were down to 68 and 65 in 1987. However, for Puerto Rican students, the distance from the White NH group has not narrowed appreciably: mean differences were 87 and 92 in 1976, compared to 87 and 89 points in 1987.

Due to recent revisions in the Student Descriptive Questionnaire, we have more information available since 1987 on factors associated with these differences. From this information, I have selected educational and demographic variables of special interest: high school grades, academic courses taken in high school, language background, and parental education, to be discussed next.

3. How do group means vary by high school grade point average and number of academic courses taken?

Insert Table 2 about here.

Table 2 shows the mean differences in SAT scores broken down by high school GPA and total years of study in six academic areas. It can be seen that with every increase in category of grade point average, there is an increase in test score mean, and this pattern is found for all groups-- White NH Latin American, Mexican-American and Puerto Rican students. For example,

if we subtract the mean for students who had an "A+" average minus the mean for those who had a "C" or worse, the Verbal and Math differences are 185 and 217 points, respectively, for White NH students. For Latin-American students these differences are 190 and 217 respectively; for Mexican-American students, these differences are 165 and 213, respectively; for Puerto Rican students, these differences are 101 and 162 points, respectively. (I compared these grade categories because they represented the bulk of students; there were very few students with grades of "D" or below -- one percent or less). These results show that test scores have a high degree of relationship to high school grades for every group.

The same pattern is found for the distribution of total years of study in six academic subjects. With each increase in the number of courses taken in high school, there is a corresponding increase in Verbal (V) and Mathematics (M) test score means, for all groups. Overall, the difference between those who took 20 or more course years vs. those who took fewer than 15 course years are 115 (V) and 128 (M) points, for White NH students, 116 (V) and 123 (M) for Latin American students, 104 (V) and 111 (M) for Mexican-American students, and 93 (V) and 113 (M) points for Puerto Rican students.

We must remind ourselves of the often repeated caution that correlation does not imply causation. The relationships shown here with course numbers cannot be interpreted causally since associations can be reciprocal. That is, students who have higher achievement levels in school will tend to take more academic courses and will have higher test means. We can also expect that students who take more courses in a given subject area will improve ir their achievement in that subject area and in related skills. Nevertheless, the results here show the pattern that we would expect to find if the tests were doing their job in terms of measuring academic skills. For every group, the

higher achieving students who take more courses tend to receive higher test scores.

Despite this constant pattern within groups, there are mean differences between White NH and Hispanic-American students when we hold constant grades or number of academic courses. The source of these differences deserves further investigation. Perhaps if parental education, language background, course grades, and number of non-remedial academic courses were controlled for jointly, (not just one at a time as they are in these tables), these differences would be further reduced. Also, as shown in the previous paper by Richard Duran (1988) for this conference, there is evidence from data sets such as High School and Beyond and the National Assessment of Educational Progress that Hispanic students are overrepresented in high schools with fewer resources or in curriculum tracks within high schools that have less demanding courses. Thus, quality of schooling may be associated with lower test scores for Hispanic students after controlling for grades, number of courses, and background variables. Unfortunately, the data set from the 1987 Profiles does not contain information on quality of the students' high school which would enable us to test this hypothesis at this time.

4. Are there ethnic differences in the distribution of numbers of academic couses taken in high school?

As shown in Table 2, there are noticeable differences between Mexican-American and White NH students in the distribution of academic courses. While 35% of the White NH population takes 20 or more year-long academic courses during high school, only 16% of Mexican-American students take this many. Twenty percent of Mexican American students take fewer than 15 course-years of academic subjects, whereas only 12% of White NH students take this few.

Unlike the distribution for the Mexican-American group, the distributions for the other two Hispanic groups resemble the White NH pattern very closely with differences smaller than 3% in every category.

5. What is the relationship between type of mathematics courses taken and test means? Are test means higher for students who take more mathematics courses?

The 1987 profiles reported breakdowns in coursework for specific subject areas in high school, from which I have selected only the mathematics courses (see Table 3) because they can be expected to have a more direct and interpretable impact on test means than do other subject areas. As explained earlier, these relationships cannot be interpreted naively to mean that taking one more course "raises" means by a specified number of points since we have two reasons why we can expect means to be higher for students who take more challenging math courses -- self-selection and honing of skills.

Insert Table 3 about here.

This is evident in Table 3 which shows the percentage breakdown and SAT means by type of mathematics courses taken for each ethnic group. Students who took trigonometry, precalculus, and calculus in high school have higher means for both SAT-V and SAT-M scores than students who took only algebra. For example, among White NH students, a little more than half the population of students take trigonometry and these score 28 points higher in SAT-V and 45 points higher in SAT-M than students who have had only algebra. For Latin-American, Mexican-American, and Puerto Rican students, the same pattern is found. Those who took trigonometry had mean SAT-V scores that were 27 to 34 points higher than those who took algebra, and had mean SAT-M scores that were

43 to 56 points higher than those who took algebra.

Also from Table 3 we can see the breakdown in means by number of mathematics courses taken. As shown in this table, students who take more years of course work in mathematics get higher scores, in both verbal and mathematics. For White NH students, SAT-V and SAT-M means differ by 73 and 174 points, respectively between the group that takes more than 4 years of math and the group that takes two to two-and-a-half years. (The groups with less than 2 years have very few cases and are not used for comparison purposes here). Respectively in the Verbal and Mathematics subtests, for Latin-American students, these differences are 87 (V) and 156 (M), for Mexican-American students, these differences are 80 (V) and 179 (M); for Puerto Rican students, these differences are 69 (V) and 132 (M) points.

The higher verbal scores among students who took trigonometry and more mathematics courses probably reflect the tendency for higher achieving students to take more challenging courses (i.e., self-selection). The increases in mathematics were larger, which may indicate that students become more skilled in applied problem solving as they take higher levels of mathematics. However, we cannot rule out that there is more self-selection for higher mathemetics skills than there is self-selection for verbal skills among students who take trigonometry and more mathematics courses. Nevertheless, the results show that students with the best preparation in mathematics get higher SAT-M means.

In the interest of brevity, I have not included here the breakdowns by numbers of courses in English, social sciences and history, art and music, foreign and classical languages, natural sciences, and computer programming or data processing. However, the patterns generally show an increase in both verbal and mathematics means as the number of courses increases. Generally,

courses in language arts and social sciences are associated with greater increases in verbal versus mathematics means, whereas the natural sciences are associated with greater increases in mathematics means versus verbal means. The relationship to SAT-V is less strong for art and music courses than for other courses in the humanities.

Now we shift the focus to variables that when held constant noticeably reduce the size of group differences.

6. How large are ethnic mean differences when we compare Hispanic and White, non-Hispanic (NH) students of the same (a) language background, (b) parental education?

Insert Table 4 about here.

As you can see from Table 4, for each aforementioned variable, there is a breakdown along two dimensions, (a) group membership (columns) and (b) values of a variable (rows). I will first look at row differences, i.e., the variation within each group according to different values of the variable. Then I will examine column differences, i.e., ethnic group differences within each value of that variable.

(a) Language Background. Table 4 shows the SAT means and percentage breakdown for each category of language background for White NH and Hispanic students. Looking at means within each ethnic group, we see that for all groups, students who learned English and another language jointly had SAT-V and SAT-M scores that were lower by 25 to 50 points than those who learned English first. The differences depend on the ethnic group and are largest for the Puerto Rican group. If we compare each ethnic group within each language category, we see a dramatic narrowing of ethnic differences when we consider

groups of the same language categories. Among students who learned English as their first language, the means for Latin American students are only 28 (V) and 34 (M) points lower than for White NH students, as compared to 60 (V) and 67 (M) points in the total group, ignoring language background. This pattern is also found for Mexican-American and Puerto Rican students and at every category of language background.

Hence, there is a very clear relationship here between language background and test performance for all groups, but it cannot be considered a strictly causal one. There are a number of other variables associated with language background, such as socioeconomic status and immigration history, that can also affect test performance.

Nevertheless, a large part of the difference between Hispanic groups and White NH students can be explained on the basis of language background and factors associated with language background. There are proportionately much fewer Hispanic students with an English-first background (21% to 44% depending on the group) in comparison to White NH students who are 94% English-first background. Thus, language background and associated factors have a much larger net impact on the overall mean for Hispanic students than on the overall mean for White NH students who have mostly learned English as their first language.

Of course there is a large body of research dating back to the classical studies by Sanchez (1932a, 1932b, 1934a, 1934b) about the intepretation of aptitude and intelligence test scores for bilingual students. Obviously, scores on an aptitude test in English are affected not only by the level of aptitude of the individual but also by his or her level of proficiency in the language of the test. This relationship was more explicitly detailed by Alderman (1982) who examined the correlations between SAT scores, English

proficiency, and aptitude in examinees' native language for Spanish-speaking students in Puerto Rico. In his study, English proficiency was measured by the Test of English as a Foreign Language and aptitude was measured by a test in Spanish used for college admissions in Puerto Rico. He found that the higher the level of English proficiency, the greater the correlation between the SAT and the aptitude test in Spanish. He partitioned the variability of SAT scores into two components (1) proficiency in English and (2) aptitude. It was clear that for students with high levels of English proficiency, the SAT variability was mostly due to aptitude and was therefore an appropriate measure of aptitude; but for students with very low levels of English proficiency, the SAT was primarily measuring proficiency in English rather than aptitude. Given that the large majority of Hispanic students living in continental United States are more proficient in English than in Spanish, this research supports the use of English aptitude test for the majority of Hispanic students. However, Alderman's findings suggest that even when English proficiency is high, there may be some extraneous variability in the aptitude test scores that is related to language proficiency, rather than underlying aptitude.

As this point we need to ask, what implications do the effects of language background have in the evaluation of test bias? But first we have to rephrase the question, because as Anastasi (1982) has pointed out, the evaluation of test validity always has to be asked in terms of the purpose and context for which the test is being used. Hence the rephrased question is: In the context of college admissions, where English is the language of instruction, do mean differences in test scores attributable to language background give unambiguous evidence that tests are biased? The answer is: no, because language background may affect college performance in the same way

20

BEST COPY AVAILABLE

and to the same degree that it affects test performance. If language affects college performance, this would imply that in a predictive sense, the tests would give an accurate reflection of college aptitude when the language of instruction is in English.

Some studies of foreign students, for example, show that the Test of English as a Foreign Language (TOEFL) has low to moderate correlations with undergraduate and graduate success in U. S. universities, although quantitative scores on the SAT or Graduate Record Examination usually have higher correlations with college or graduate grades than does the TOEFL (Hale, Stansfield, & Duran, 1984, reviews of studies #4, 11, 70, 71, 72, 78). Thus, there is evidence that for foreign students, proficiency in English does have a modest impact on college performance in the majority of cases.

Of course these results found with foreign students may not be generalizable to native-born American students with bilingual or multilingual backgrounds who can be expected to have a much higher level of English proficiency. There are two interesting empirical questions that need to be addressed by examining language background and college performance. Specifically, (1) do bilingual students perform better in college than one would expect on the basis of their test performance? and (2) is there more accuracy of prediction (a higher validity coefficient) for students who are monolingual English speakers in comparison to the accuracy for bilinguals? Currently there is an ongoing College Board study of the extent to which language variables affect the predictive validity of the SAT. We should have preliminary results for this study within the next six months.

(b) Parental education. Table 4 shows the percentage breakdown and SAT means by levels of parental education. For White NH and Latin American students, the mean SAT-V and SAT-M scores for the group whose parents did not

complete high school are approximately 100 points lower than for those whose parents had graduate degrees. These differences are 89 (V) and 74 (M) points for Mexican-American students and 74 (V) and 97 (M) points for Puerto Rican students. For every group, average test scores increase with parental education level.

Looking within each category of parental education level and comparing ethnic group means, we see a narrowing of ethnic group differences when parental education is the same. Latin Americans differ from White NH students of the same parental education by 35 to 56 points for the SAT-V test, and 31 to 47 points on the SAT-M test. These differences are smaller than for overall group differences of 60 and 57 points when groups are not broken down by educational level. A similar pattern is found for Mexican-American students where the differences range from 36 to 60 points for the SAT-V test and 29 to 59 points for the SAT-M test, when groups of the same parental education levels are compared. These differences are smaller than the overall group differences of 68 (V) and 65 (M) points when parental education level is ignored. For Puerto Rican students, the differences range from 61 to 88 points for the SAT-V test and 70 to 77 points for the SAT-M test. These differences are smaller than the overall group differences of 87 (V) and 89 (M) points when collapsing across parental education levels.

Hence, there is evidence here that ethnic group differences are associated in part with parental education level, and these differences are reduced when parental education levels are the same. Again, this finding is not surprising given the extensive body of evidence showing that students with well educated parents receive a higher quality of education at home and attend better schools. Therefore, they are generally better prepared for college, and this advantage can be expected to be reflected in higher test scores.

Evaluations of Test Item Content

A relatively new methodology for investigating possible test bias at the item level has emerged in the last 15 years. Typically what is done is to contrast members of subpopulation groups (e.g., male vs. female, Black vs. White examinees) on their performance on individual items after controlling for overall total test score. Research in this area identifies items that are relatively easier or harder for one group vs. another, after taking into account overall score. One way of categorizing this research is to say that it identifies items that are inconsistent with total test score for one subpopulation.

Initially, the area of study was called "item bias" research, but is now more often referred to as the study of "differential item difficulty" or "differential item functioning," abbreviated DIF. The change in terminology arose from the many instances in which items found to have DIF were not necessarily biased or unfair. The judgment about an items' fairness is usually made on the relevance of the item to the trait being measured, what psychometricians call content validity.

An example of an item that had DIF but was not considered biased was reported by Breland, Stocking, Pinchak, & Abrams (1974). They found that an item requiring familiarity with square roots on a mathematics achievement test was relatively more difficult for Hispanic, Black, and American Indian students than for White NH students. Since the source of the discrepancy reflected a real deficiency in the students' knowledge of basic concepts necessary for success in mathematics courses, it cannot be considered "unfair" or "biased." However, if a similar discrepancy relating to square roots was found for an item on a reasoning test, one could argue that specific knowledge of square roots was independent of reasoning ability and was therefore

introducing extraneous sources of difficulty in the reasoning test. Hence, an item on a reasoning test involving square roots and demonstrating DIF would be considered biased.

In addition, it is possible that items or aspects of the test format that have references or language that might be stereotypical or objectionable to a subgroup may not necessarily show statistical discrepancies. Nevertheless, the consensus among psychometricians is that such items are unfair and should be eliminated, even if there is no evidence of an increase in performance for members of that group when the item is deleted (Shepard, 1982). It is standard procedure that test items at ETS undergo a sensitivity review by a panel of judges before being included in a test form (see Hunter & Slaughter, 1980). Objectionable items are eliminated or modified before inclusion in a test form.

Items that survive these judgmental reviews are administered to examinees and are later analyzed statistically as part of preliminary item analyses before reporting scores, to see if they show DIF for ethnic and gender subgroups. Items that are determined to be statistically discrepant for certain groups (i.e., those with large statistical indexes of DIF) are flagged for scrutiny by panels of judges to identify the sources of group differences in performance. If the source of the difficulty is judged to be irrelevant to the test specifications, then the item is not included in computing the score. Also, pretested experimental items that show DIF are usually modified or eliminated from the pool of items to be used in assembling future tests. I know from personal experience in serving on two of these panels that occasionally, these statistical methods catch subtle, unexpected content effects that get overlooked by sensitivity reviewers. Thus, the statistical DIF analysis procedures lead to refinements in the test.

Two of the most frequently used statistical methods with college admissions tests are the Mantel-Haenszel and the standardization methods. These procedures subdivide members of the two groups according to intervals of total test score. Then individuals from the two groups at the same score level are compared with respect to their performance on the given item. Although the specific statistical index values for the two methods are on different scales, they are almost perfectly correlated and classify items in the same way.

In the early years of DIF research, judges were not very successful in predicting which items would be discrepant, or in finding reasons to explain group differences for those items that turned out to be statistically discrepant for certain groups. Often, items judged to be objectionable or differentially more difficult on an a priori basis did not show any group differences statistically, and judges often disagreed with one another (see review by Pennock-Roman, 1986, pp. 202-203). Now that we have more accurate statistical methods, the evidence about what content characteristics of items tend to produce DIF has been more consistent and interpretable. Often it is possible to formulate hypothesis about the characteristics of items that lead to DIF and these predictions are frequently confirmed with results from another study.

For example, one of the most consistent findings for Hispanic students, both foreign and American-born, is that vocabulary words that are true Spanish cognates are relatively easier for Hispanic examinees than for White NH examinees (Breland et al., 1974; Alderman & Holland, 1981; Chen & Henning, 1985; Schmitt, 1988).

The study by Schmitt (1988) deserves special attention because it is the most extensive analysis to date of item characteristics associated with DIF

BEST COPY AVAILABLE

for Hispanic students. She used the standardization method which compares the percentage of Hispanic vs. non-Hispanic White examiness answering the item correctly, when controlling for total score. Before discussing her findings, it is important to understand what this index means and what it does not mean.

The cutoff value for the statistical index for the standardization method (D_{STD}) was set at .05 for this study. When the value of the index for a particular item exceeds a value of .05, it means that on the average, Hispanic examinees answer the item correctly (or incorrectly) 5% more often than White NH students with comparable scores, which is a very small difference in performance between the two groups. It does not mean that all of the members of one group failed it and that members of the other group answered it incorrectly. Furthermore, each flagged item is independent of performance on other flagged items. An examinee that correctly answers one of the differentially easier items for his or her group will not necessarily answer correctly all of the other items that favor his or her group. Because the effects found by the statistical procedures are subtle and the responses of individual members of a group to the set of flagged items vary, the overall effect on total score produced by flagged items tends to be very small. For example, Shepard, Camilli, and Williams (1986) found that eliminating flagged items on a test changed group means on the test only by a trivial amount.

Schmitt examined items that showed DIF in two alternate administrations of the SAT in 1983 and 1984 (which she calls Study 1 and Study 2) with very large samples. The cutoff set for this study was lower than the usual cutoff that is used to flag items for scrutiny[1] in operational procedures for generating scores or assembling tests. This lower cutoff was set for research purposes in order to cast a wider net and have a larger number of potentially discrepant items that may show group differences. The SAT-M showed few items

with DIF in both studies, so that her descriptions focused on the verbal subtest. From Study 1, she identified four characteristics that were apparently associated with DIF on the verbal test. Then she rated the items for the SAT form in Study 2 to see how strongly these characteristics were associated with DIF results.

The factors she identified that were associated with higher performance for Hispanic vs. White NH students in both studies after controlling for overall test scores were:

(1) True cognates, or words with a common root and common meaning in English and Spanish (e.g., pallid and pálido). There were some exceptions, which the author attributed to the presence of other elements of the item that cancelled this effect.

(2) Reading comprehension items of special interest to Hispanic students. Specifically, in Study 1, a passage on Mexican-American women was relatively easier for Mexican-American students, but not for Puerto Rican students of the same overall score level. In Study 2, a passage on a Black mathematician was relatively easier for both Puerto Rican and Mexican-American groups in comparison to White NH students of the same overall score level.

Factors associated with lower performance for Hispanic vs. White NH students in both studies after controlling for overall test scores were:

(1) False cognates, i.e., words that look identical or similar in the two languages but have different meanings in the context of the item (such as "enviable", which in Spanish means capable of being mailed, or transportable). It should be noted that a given pair of similar words in the two languages can be true cognates for one item and

false cognates on another because words have multiple meanings. Some of the meanings match in both languages but others do not.

(2) Homographs, or words that have more than one meaning (e.g. bark of tree and bark of a dog).

(3) In another analysis (Schmitt & Dorans, 1987) found that vertical relationships (which are word associations extraneous to the analogical relationship) between the stem and key or the stem and distractors in analogy test items also tended to handicap the performance of Hispanic examinees.

The effects related to language were more strongly evident in the Puerto Rican group in both studies, because there is a greater incidence of bilingualism among college bound Puerto Ricans than among college bound Mexican American students (see Table 4 in the section on language background and mean differences). Given that the reading passages provide more context for responses, it is not surprising that the majority of the items that handicapped Hispanic students were found in the antonym and analogy sections of the test.

These findings are highly consistent with research on DIF for Black examinees which has also shown some of the same characteristics as those found by Schmitt. The greater proportion of flagged items has been found among antonym and analogy items, (Dorans & Kulick, 1987; Rogers & Kulick, 1987; Schmitt & Bleistein 1987; Freedle & Kostin, 1987). Complicating the explanation of findings is that Black and Hispanic examinees tend to reach fewer items than White NH students with the same total score. Researchers found it difficult to disentagle factors that appeared to be associated with DIF because many characteristics were confounded with item position. However, when this differential speededness effect was controlled, there were still

proportionately more analogy items that were flagged as discrepant. Generally, minority examinees performed relatively better on more abstract, more suposedly difficult antonyms and analogies occurring later in the section, and worse on early, easier items that had homographs. Furthermore, "vertical relationships" or extraneous associations between the stem and distractors also tended to handicap Black and Hispanic examinees. A think-aloud procedure with 11 Black and 11 White NH students suggested that Black students do relatively better with more abstract, difficult analogies than on the easier ones, and this effect was found independently of item position (Freedle, Kostin, & Schwartz, 1987). However, this result is based on few cases and needs to be verified in future studies.

Currently, there are two ongoing studies by researchers at ETS in which item characteristics were experimentally manipulated to see their effects on DIF results (Scheuneman, personal communication January 1988; Schmitt, personal communication February 1988). In the next year, we will have more solid information on these issues.

It is important to note that proportionately very few items showed DIF that handicapped Hispanic students on the SAT-V in these two studies. As shown in Table 1 from Schmitt's study, out of a total of 85 items there were only 5 discrepant items that handicapped Mexican-Americans and 7 for Puerto Ricans in Study 1. These items were partially counterbalanced by 3 other items that favored Mexican-Americans and 5 that favored Puerto Ricans. For Study 2, (reported in Table 3 of Schmitt's study), out of a total of 85 items, there were 10 items that handicapped Mexican-Americans and 9 that handicapped Puerto Ricans. These were partially counterbalanced by 4 items that favored Mexican-Americans and 7 that favored Puerto Ricans.

Although Schmitt did not analyze how much the mean difference between Hispanic and White NH students could be reduced if the discrepant items were eliminated, it is unlikely that discarding these items would have had much effect on total scores. The items handicapping Hispanic students were relatively few in number and were partially counterbalanced by the items that favored Hispanic students. Furthermore, the statistical discrepancies were small[1]. The index of the D_{STD} had values that equalled or exceeded .11 only once for 85 items in Study 1 and only once in 85 items for Study 2, and in both cases the most discrepant item favored Hispanic examinees. This means that the differences between groups in the probability of correctly answering the items were noticeable but not large enough to make an enormous difference in total score. Thus, it is very unlikely that eliminating these items would have substantially reduced ethnic differences on the test.

Another study that led to interpretable DIF results on Hispanic students with the American College Test (ACT) was reported by Loyd (1982). She found that reading passages in the English Usage test had six discrepant items, three favoring White NH students and three favoring Hispanic students. Two of the items favoring White NH students had interpretable results. She found that these items involved skill in punctuating or adequately placing adjectives and adverbs in a series. These are linguistic features that may have been more difficult for bilingual students. In the Social Sciences Reading test, there were seven discrepant items, three favoring Hispanic students and four favoring White NH students. Two out of four of the items favoring White NH students required knowledge of the subject matter that was not contained in the reading passage. Thus, the latter finding suggests a deficiency in the educational background of the Hispanic candidates that made these items relatively more difficult. As with the SAT findings, it is

unlikely that overall mean difference could be substantially reduced if the flagged items were deleted. There were relatively few flagged items and about half of them favored Hispanic students.

Although the small number of flagged items probably cannot completely account for mean differences between Hispanic and White NH students, the results provide important information about the effects of bilingualism and other factors on test performance. It appears that bilingualism has both advantages and disadvantages in test performance, depending on particular linguistic features of the items. In discussing Schmitt's findings at a conference, Shepard (1986) proposed that false cognates and homographs introduce irrelevant sources of difficulty but that true cognates are not necessarily unfair. She recommended that "the proportion of Latin roots [items in the test should] mirror what is found, say, in typical freshman reading assignments." (p. 3). However, by the same reasoning, one can argue that false cognates and homographs also occur in college texts and that they, too, should be proportionately sampled.

Thus, it is evident that procedures that detect discrepant items serve a very important function in revealing test content characteristics that give unexpected results for some groups. They are essential to opening a discussion about what type of content should be specified in a test. Shepard (1986, p. 5) has pointed out that ideally, these methods can help us "to search out sources of irrelevant difficulty and to arrive at a better understanding of what a test measures." They can also serve an important diagnostic function because they can point to gaps in minority students' backgrounds. For example, in Breland, et al. (1974), an item on square roots in the mathematics was found to be relatively more difficult for minority

students. Thus it revealed a content deficiency in students' backgrounds that could be used to design curriculum for remedial instruction.

Nevertheless, it is important to keep in mind that this type of methodology is limited because results are always relative to other items on the test. Since the statistical methods control for overall test score to identify discrepant items, they cannot tell us if the test score as a whole is artifactually depressed for one group. To test whether the test score as a whole is biased is best achieved by an analysis of the accuracy of prediction of test scores for identifying who will succeed in college for different groups.

Predictive Validity of Admissions Tests for Hispanic Students

As we have seen in the previous sections, (1) Hispanic students score substantially below White, non-Hispanic (NH) students on selective admissions tests and (2) there are certain kinds of items that are differentially easier or harder for Hispanic students. In terms of formulating policy, it is important for research to determine whether these differences are reflected in college performance. One of the most direct ways that we can determine the accuracy of the information that tests provide about college aptitude for Hispanic students is to examine how well tests predict college grades. Unfortunately, predictive validity studies are hampered by many practical difficulties.

Methodological Difficulties in Predictive Validity Research. Investigations on predictive validity for Hispanic students have been few in number because there are many practical problems in obtaining large sample sizes in selective colleges. First, the sensitivity (power) of statistical methods to detect differential validity is reduced when sample sizes for

BEST COPY AVAILABLE

Hispanic groups are small. Second, there are also problems in securing adequate identification of which students are truly Hispanic. For example, sometimes Spanish surname has been used as the only identifier for Hispanicity. Census data indicate that Spanish surname fails to identify as Hispanic about a third of students who consider themselves Hispanic. Furthermore, about a third of persons with surnames judged to be Spanish do not consider themselves Hispanic because they may be of Italian or Portuguese heritage or have only one very remote Spanish ancestor whose name has been handed down several generations to persons of mostly non-Spanish heritage. When self-reported ethnicity is used, we find that it is sometimes incomplete or inaccurate at many institutions.

Third, validity studies should include several institutions because results at one university may not be generalizable to others. The amount of selectivity of an institution reduces the variance in the predictors which decreases the values of correlation coefficients and other indices of prediction. Institutions with high variability among their students generally show higher correlations between college grades and test scores. Even when we control for differences in selectivity, there are variations in grading standards within an institution and between institutions that affect how well tests can predict performance at any given university. In sum, evaluations of tests should involve many institutions of different types.

Fifth, college grade point average (GPA) -- which is the usual criterion of college success against which tests are evaluated -- has many limitations. Grades are internally inconsistent (unreliable) because they vary unsystematically from instructor to instructor and also vary systematically across different fields of study. For example, Strenta and Elliott (1987) have documented that some departments, especially engineering and the physical

sciences, have much harsher grading standards than others. The difficulty levels of individual courses are not taken into account, despite the fact that an "A" in a remedial course does not mean the same thing as an "A" in an honors course. If Hispanic students take more remedial courses or more science courses than do White NH students, then their college GPAs are not comparable and it presents a serious problem in doing a validity study because artifactual effects will be found. These problems limit the reliability and validity of grades as a measure of college success, thus artifactually lowering correlations between grades and other measures.

A Review of Regression Terms Used to Test Predictive Accuracy in Two Groups. In comparing the accuracy with which tests predict college grades in a majority or reference group vs. a minority or focal group, the preferred method is the use of regression equation equations. This statistical procedure yields several indexes of interest. One is the multiple R, which measures the overall accuracy of prediction of the college grades using all of the predictors. (If there is only one predictor, the multiple R and the Pearson correlation coefficient are the same.) When the multiple R is squared, it gives the proportion of variance in the college grades that is explained by the predictor variables. This index is free of the units of college grades. Unfortunately, the multiple R index is subject to some artifactual effects. If the variability of college grades, high school grades, or test scores is restricted in one group and less restricted in another, the multiple R, like a zero-order correlation coefficient, can appear to be lower in the group with restricted variance even when the groups differ little in accuracy of prediction.

Because of the artifactual problems involved in interpreting multiple Rs, there is another index that is preferred for comparing groups -- the standard

error of estimate. The standard error of estimate is the amount of variability of the residuals which are the differences between actual and predicted college grades. It measures the amount of scatter of points away from the regression line and is a function of both the multiple R and the variance in college grades. The larger the multiple R, the smaller the scatter away from the regression line and the smaller the standard error of estimate. However, unlike the multiple R, the standard error of estimate is in the same units as the original college grades; and it is less subject to interpretation problems if there is restriction in the variance of college grades, because these restrictions in variance are also reflected in the standard error of estimate (see Cohen & Cohen, 1983, p. 104).

In applying regression to detect group differences, we would expect the multiple R to be smaller (less relationship) and the standard error of estimate to be larger (because there would be more scatter) for Hispanic students if tests were less accurate in predicting college grades for Hispanic students than for White NH students.

A third index important in multiple regression is the regression weight for each predictor. In the regression equation, estimates of individuals' college grades are found by weighting their HSGPAs and test scores and summing these weighted values. The more that a variable contributes to prediction independently of the other variables, the higher its regression weight. Thus, these weights depend on the other variables in the equation. For example, when test scores are the only predictors, their regression weights are larger than when HSGPA is included as a predictor together with test scores. If differential validity exists, and tests are relatively better predictors for White NH students than they are for Hispanic students, we would expect that the regression weights for White NH students' test scores would be larger.

A fourth index of interest is the regression intercept, which represents the point at which the regression line crosses the axis of college grade values (Y) when all of the predictors have zero values. When there are two groups to be compared that have equal regression weights for the predictors, the difference in their intercepts reflects differences in the average value of college grades for any given value of the predictors. That is, if Hispanic students were to get higher grades in college than do White NH students with the same test scores despite having equal weights for the predictors, this means that the Hispanic students' regression line as a whole is higher on the graph than the line for White NH students; in other words, the intercept is higher for Hispanic students than it is for White NH students. If the intercept were higher for Hispanic students, assuming equal regression weights, one would expect that applying the White NH students' regression line to values for Hispanic students would underpredict Hispanic students' actual college performance. This underprediction can also occur if there are group differences in regression weights such that some portion of the regression line for Hispanic examinees is higher on the graph than the regression line for White NH examinees.

In addition, some researchers have also examined how much improvement in the accuracy of prediction is achieved when tests are added to high school grades, in comparison to the accuracy found when high school grades are the sole predictor (this is called incremental validity of tests). It involves taking the difference in two multiple Rs, the multiple R when the equation includes HSGPA plus test scores minus the multiple R when only HSGPA is used. The difference in multiple Rs measures the improvement in selection of students for admissions (Beaton & Barone, 1981) whereas the difference between the two multiple R^2s measures the amount of additional variance in the college

grades that is explained by adding test scores. Hence if tests count relatively more for White NH students, the incremental validity of the tests should be higher for White NH students than for Hispanic students.

To summarize, consider what we would expect to find among these indexes if tests were not measuring college aptitude among minority students as well as they do among majority students. First, we might expect that the degree of relationship between scores and college grades would be smaller for Hispanic students, leading to smaller multiple R's and larger standard errors of estimates for Hispanic students. A second way that tests could be biased is if the verbal or mathematics sections or both subtest scores counted less as predictors, leading to lower regression weights for test scores in the regression line for Hispanic students than in the line for White NH students. A third possibility is that test scores systematically underpredicted the college performance of Hispanic students -- that is students would receive higher college grades than one would expect on the basis of test scores, which could arise in several ways due to differences in intercept values or regression weights for the two groups. A fourth possibility is that if the tests counted less in the prediction of college grades for Hispanic students, the improvement in prediction when tests are added to grades (difference in multiple Rs as defined above) would be smaller for Hispanic students than for White NH students.

Thus, there are several basic questions that are generally asked in comparing regression lines for two groups. One question is how strong is the overall relationship of all predictors (taken jointly) with the criterion, (which is measured by the standard error of estimate and the multiple R). A second question is whether there are differences in the degree of relationship between each predictor and college grades (the raw regression weight for each

variable). The third question is whether the use of the reference groups' regression equation systematically overpredicts or underpredicts grades for most persons in the minority or focal group (which could be the result of differences in regression intercepts and/or regression weights). The fourth question (which is not always asked) is whether the incremental validity of tests beyond the prediction accuracy found with just HSGPA is lower for Hispanic students than for White NH students.

Duran's Review (1983) of Predictive Validity Studies. The most complete review of predictive validity studies for Hispanic students was authored by Duran (1983), in which more than 14 independent analyses were reviewed. The general findings were that:

(1) Overall, there were no dramatic differences in regression systems between Hispanic and White NH students, although some subtle differences were consistently found. In general, Hispanic students tended to perform less well in college than did White NH students, to a degree that was commensurate with their high school grades or rank and lower test scores.

(2) The most consistent subtle difference found was that often there were lower multiple Rs for Hispanic students, for all predictors but especially test scores. These differences tended to be small and non-significant. Rarely did the authors of the validity study discuss whether differences in the multiple Rs and correlations were due to differences in group variances for predictors and college grades In a footnote Duran cautioned (1983, p. 139) that "a more sensitive analysis of differences in prediction should rely on interpretation of standard-error-of-estimate statistics. For the most part, standard-error-of-estimate statistics were

not directly available in the predictive validity studies reviewed in this report; in contrast, multiple $\underline{R}$ or $\underline{R}^2$ statistics were readily available for studies."

(3) The median zero-order correlations between college grades and predictors showed that the highest correlation found for Hispanic students was for HSGPA and it differed little from the correlation found for White NH students. In contrast, the correlation with quantitative scores was the lowest and had the largest difference between Hispanic and White NH students. This correlation was reported separately for only 9 studies.

(4) Few studies reported explicitly the incremental validity of tests over the prediction achieved with HSGPA. Goldman & Widawsky (1976) found that it was less than 10% at four campuses of the University of California for Hispanic students. They explained this finding by pointing to a larger correlation between test scores and high school grades for Hispanic students than the correlation for White NH students.

(5) None of the researchers who tested for ethnic differences in regression intercepts found evidence of underprediction of Hispanic students' grades, and in fact one study found substantial overprediction (that is Hispanic student's actual college performance was lower than that predicted by the White NH equation). But many researchers did not explicitly test for under- or over-prediction.

Hence, the evidence thus far shows some subtle evidence of lower accuracy of prediction of tests for Hispanic students in comparison to White NH students. Nevertheless, tests have some incremental validity over the prediction based on grades alone for Hispanic students, and this incremental

validity varies according to university. For example, it was substantial for students at the University of California at Davis in the Goldman & Widawski (1976) report. These conclusions must be considered tentative because there are many limitations in the studies reviewed thus far. These limitations include the following:

(1) Because most of the universities studied were public institutions in the southwest, the research is primarily based on Mexican-Americans; no data on Latin American and Puerto Rican groups are available to date on predictive validity. The states and types of institutions sampled are also limited.

(2) Often, Hispanic students were only one of several racial/ethnic groups considered so that results were not reported completely enough to examine all the questions we would want answered. Frequently important information such as intercorrelations, standard errors of estimate, and degrees of incremental validity were not reported.

(3) Although some studies did take gender into account by doing regression analyses separately for males and females, this control was not available in all studies. Controlling for gender can make a difference if there are relatively more females in one group than in another because females consistently get higher college grades than males for the same level of test scores. Furthermore, very few studies controlled for the effect of different majors on college grades.

(4) The effects of language on predictive validity were not addressed by the majority of studies.

Currently, I am directing a study funded by the College Board to address most of these issues and the results will be forthcoming soon. (A report on differential validity will be reviewed in May 1988 by the College Board). This study includes six universities, three public and three private. Two universities in the northeast have Hispanic students that are predominantly Puerto Rican, and one in Florida has students that are mostly Cuban American. In the analyses, we examine effects of gender and major on college grades, and we look at incremental validity and standard errors of estimate. In a second report we will have additional detailed results on the effect of language background on college performance for Hispanic students who are bilingual.

Student Awareness of Types of Preparation Needed for College and Admissions Tests

When we focus on ethnic differences in admissions test scores and how lower scores affect access to college for Hispanic students, we tend to overlook what may be the largest problem for access -- the fact that so many Hispanic students do not take admissions tests at all. Hispanic students are proportionately overrepresented at two-year colleges that traditionally do not require taking an admissions test and are underrepresented in the population that seeks acceptance to four-year-institutions. A study by Lee and Ekstrom (1987) has shed light on some of the complex factors that influence the flow of students in the educational pipeline. In the abstract and discussion, they summarized the findings of their study as follows:

> Using data from the first and second follow-ups of High School and Beyond, including student self-reports test scores, and high school transcripts, we found that guidance counseling services appear to be unequally available to all public high school students. Students from families of lower socioeconomic status (SES), of minority status, and from small schools in rural areas are less likely to have access to guidance counseling for making .... important decisions [about selecting a curriculum track or planning an appropriate course

> of study] at the beginning of their high school careers. Moreover, students who lack access to guidance counseling are more likely to be placed in nonacademic curricular tracks and to take fewer academic math courses. It appears that students who may need such guidance the most, since they come form home environments where knowledge of the consequences of curricular choices is limited, are least likely to receive it in their schools. (p. 287)
>
> [Specifically,] less than one-fourth of all high school students select a curriculum with any assistance from a counselor, and only about half of all high school students receive counselor assistance in program planning. Moreover, only slightly more than half of all high school students have access to counseling for their plans after high school.... These figures suggest that there is likely to be a group of students who might have either the ambition or the ability to attend college but who have no contact with a counselor until the end of their high school years. As a consequence, such students may not have entered a curriculum track providing preparation for college or, regardless of track placement, may not have taken courses that are either necessary or desirable preparation for college." (p. 306).

Although Lee and Ekstrom (1987) do not explicitly address how counselor access may affect students' preparation for taking college admissions tests, we can certainly expect that these inequities in access to counseling lead to a lack of information about how students should prepare for admissions tests. This lack of guidance probably exacerbates ethnic group differences in test scores because Hispanic students are overrepresented in schools with poor resources.

Fortunately, some states such as California are addressing this problem. As a result of the Tanner initiative, a recent program in California has been implemented to provide minority students in disadvantaged inner city and rural districts with more college admissions counseling and test-taking guidance. This program reaches out to many students who would normally not attempt to take the SAT and who would most likely not be admitted to four-year colleges.

Concurrently, in a collaboration between Hispanic Higher Education Coalition, ETS, and the College Board, a kit to help students prepare for the Preliminary Scholastic Aptitude Test (PSAT) was developed (College Board,

1988). This kit, Preparing for the PSAT/NMSQT for Hispanic High School Students (in press [2]) was developed primarily by Lorraine Gaire with initial assistance from Charlene Rivera. It was designed to encourage more Hispanic students to register and take the PSAT and become better prepared for college. It has enough material for a fairly lengthy (one-semester or more) orientation program and includes review of basic mathematics courses, and many other lesson plans. This kit was piloted at several school districts that have implemented the Tanner Act program.

Don Powers, Monte Perez, and I recently surveyed (October 1987) student participants (mostly 9th, 10th, and 11th graders) in several of these programs and obtained their reactions to the kit. The students' reactions to the test-familiarization kit were overwhelmingly positive. It was apparent from their comments that they viewed the course as an opportunity to improve their problem-solving and basic skills not just to gain test-wiseness. Most of the students wanted the program to be extended and to have more materials. As a result of the program, the number of students intending to take the PSAT or SAT increased from 58% to 86%.

Thus, it is important to note that the group of students involved in the survey included 42% of students who most likely would not have attempted to take the PSAT or SAT tests, and thus these results give us a window on students who are normally not included in our SAT samples. The results of the survey revealed the general neglect these students experience in guidance about the college admissions process and test preparation. One student commented that before participating in the program, he was not aware that admissions tests were required for admission to many colleges.

Before viewing the survey results, I expected much of the material in the kit to be new to the students, but I expected perhaps 95% to be aware of how

to fill in the answer sheet, and perhaps 80% to be familiar with test directions, test-taking strategies and the more common types of items such as reading comprehension. But I was wrong. More than 45% of the students found that they learned something new about the unit on answer sheets, more than 90% learned something new about budgeting their time, understanding the PSAT directions, when and how to guess, and how to approach different kinds of test questions. One student commented that before the program, she didn't know how to tell the difference between antonyms and synonyms. This lack of awareness about routine test-taking skills is surprising, given that the use of multiple choice tests is so widespread. It suggests that in the school districts represented in the survey, multiple-choice tests are administered without adequate preparation of the students, and that insufficient time is dedicated to a diagnostic review once test results are received.

In sum, together with Lee and Ekstrom's (1987) findings, our experience suggests that there are many Hispanic students with the ambition and motivation to attend college who lack even the most basic guidance information on how to prepare themselves for college and for admissions tests. This is a population that usually does not appear on tables of results on the SAT. They are the ones who have the most barriers to access to college, because they find out too late what steps to take for the college admissions process. The successful implementation of the Tanner Act programs suggests that comprehensive guidance counseling and test-familiarization can make a big difference in these students' lives.

BEST COPY AVAILABLE

Summary and Conclusions

Let us recapitulate some of the points made in this review.

(1) Mean differences on the SAT between Hispanic and White, non-Hispanic (NH) students are relatively large, particularly for Puerto Rican students, and they are associated with differences in language background, parental education, high school grades, and type of academic courses taken. The relationship between test scores and the aforementioned factors are consistent with the view that the tests measure the quality of a students' preparation for college in which the language of instruction is English. However, predictive validity studies are the only way to evaluate whether the mean differences in test scores reveal real deficits in the quality of preparation for college.

(2) In studies of differential item functioning or DIF, the numbers of items showing differential difficulty levels have constituted only a small percentage of the total test items, and the results have not been linked to differences in predictive validity. Hence, it is not known if the characteristics leading to unexpected group differences in items represent irrelevant sources of difficulty or if they correspond to real differences in college performance. Some kinds of test item types (specifically, analogies and antonyms) tend to be differentially more difficult for Hispanic students and other minorities. There are some indications that the problems occur primarily with the supposedly easier test items, perhaps because they have more homographs. It is interesting that some results suggest that more abstract kinds of relationships and words are relatively easier for minority students. For Hispanic students, bilingualism is sometimes an asset and sometimes a handicap. Items that contain English words that are true cognates

of Spanish words in the stem and answer choices are easier, and those with false cognates are more difficult. Reading passages with content of special interest for minority students are also relatively easier for minority students.

(3) The evidence on predictive validity suggests that tests are slightly less accurate in predicting Hispanic students' success in college than they are for White NH students, but this conclusion is based on limited evidence. More research needs to be done to investigate more fully why the correlations are lower. In particular the effect of language factors and artifactual effects of course difficulty on grading standards need to be investigated. In the majority of studies, there was no evidence that the tests underestimated the college performance of Hispanic students.

(4) The largest barrier for access to college for Hispanic students may be inequity in the availability of guidance counseling in junior and senior high school. Since Hispanic students' parents are often not college educated, their family resources cannot compensate for this lack of adequate guidance. Many students with the desire to attend college receive little or no orientation and thus enter non-academic tracks, or take the wrong courses, and fail to get basic information about college admissions and test preparation. There may be a very large proportion of Hispanic students who inadvertently avoid taking the SAT or the ACT, not realizing their connection to college admissions.

The evidence concerning the adequacy of college admissions tests for Hispanic students is correlational and not experimental in nature and as such has many ambiguities and missing information. Based on the data that are available at this time, I believe that there is room for improvement in

current admissions tests, but the major cause of differences in tests scores between Hispanic and White NH is inequity in quality of schooling and guidance counseling. Some subtle effects related to item formats and types of wording have been found by DIF research, but these effects are too small and infrequent to account for the large gaps in means.

On the other hand, as you have seen, there are some large mean differences between students who have had 15 vs. 20 academic course years in high school, and specific kinds of courses in mathematics. As pointed out by Messick (1981) coaching or short-term study for test preparation tends to lead to negligible score gains because underlying skills are not sufficiently altered. However, his analysis suggests that large score gains on the SAT can be achieved with long-term preparation (at least one semester long) that develop the overall educational skills and background of the student. Thus we can expect that the best way to raise the mean scores for Hispanic students is to ensure that they enter academic tracks in school and take as many challenging courses as they can fit into their schedule, beginning in ninth grade, if not sooner.

This step is particularly crucial for Mexican-American students. In looking at the course-taking patterns of students who have taken the SAT shown in the first part of this paper, it appears that Mexican American students, the largest Hispanic group, are not taking adequate numbers of courses for preparation in academic areas, and we can expect that the situation is much worse if we were to include all of the other students who do not attempt to take the SAT or ACT.

For many decades validity research has shown that high school records are better predictors of college performance than aptitude test scores. However, aptitude test scores give an objective basis for correcting for differences in

competitiveness among high schools. A "B" from a magnet school attended by the best students in a school district is not the same as a "B" from a less competitive high school. Thus, the value of adding the SAT or ACT to high school grades for college admissions in the evaluation of Hispanic students depends in part on the university. If a university draws students from a very heterogeneous collection of high schools, test scores can help admissions officers to evaluate student records. The more selective a university is, the more relevant it finds the test score information, because the applicants to highly selective institutions are often mostly A-average or B-average students. The test score information helps to identify who received the better quality of preparation and can keep up with the pace of work at that institution. Although the few studies so far suggest this corrective function served by tests is more successful for White NH students, each university has to conduct its own evaluation of the incremental validity of tests for Hispanic students in their own circumstances, in order to make the best use of test score information.

Future research should address the following questions: Does the accuracy of prediction of college grades decrease or increase when test items that are differentially harder or easier for minority students are included? Do tests underestimate the college performance of bilingual students? How accurately do test scores predict grades in college for Hispanic students when differences in grading standards by fields and course difficulty are taken into account? How can we improve access to adequate counseling and college preparation for minority students?

Furthermore, these questions need to be investigated with a wider variety of tests. Currently we mostly have information on aptitude tests for undergraduate admissions.

Returning to Linn's (1982) caution cited at the beginning of this paper, we must keep in mind that the psychometric quality of tests is only one component in the evaluation of tests; the benefits and losses that using tests can potentially bring to institutions, individuals, and society as whole must also be considered.

Selected Bibliography

Alderman, D. L. (1982). Language proficiency as a moderator variable in testing academic aptitude. Journal of Educational Psychology, 74, 580-587. (Also TOEFL Research Report #10, 1981, ETS Research Rep. no. 81-411. Princeton, NJ: Educational Testing Service.)

Alderman, D. L. & Holland, P. W. (1981). Item performance across native language groups on the Test of English as a Foreign Language. TOEFL Research Report 9. Princeton, NJ: Educational Testing Service.

Anastasi, A. (1982) Psychological Testing. New York: Macmillan.

Beaton, A. E. & Barone, J. L. (1981) The usefulness of selection tests in college admissions. ETS Research Report no. RR-81-12. Princeton, NJ: Educational Testing Service.

Breland, H.M. (1979). Population validity and college entrance measures. Research Monograph No. 8. NY: College Entrance Examination Board (CEEB).

Breland, H.M., Stocking, M., Pinchak, B. M., & Abrams, N. (1974). The cross-cultural stability of mental test items: An investigation of response patterns for ten sociocultural groups. ETS Project Report 74-2, Princeton, NJ.

Chen, Z. & Henning, G. (1985). Linguistic and clutural bias in language proficiency tests. Language Testing, 2, 155-163.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. (2nd. ed.) Hillsdale, NJ: Lawrence Erlbaum Assoc.

Cronbach, L. J. (1975). Five decades of public controversy over mental testing. American Psychologist, 30, 1-14.

Dorans, N. J. & Kulick, (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. Journal of Educational Measurement, 23, 4, 355-368.

Duran, R. P. (1983). Hispanics' education and background: Predictors of college achievement. NY: CEEB.

Duran, R.P., Enright, M.K., & Rock, D.A. (1985). Language factors and Hispanic freshmen's student profile. College Board Report No. 85-3, NY: CEEB .

Freedle, R. & Kostin, I. (1987). Semantic and structural factors affecting the performance of Black and White examinees matched foı ability level on analogy items from the Scholastic Aptitude Test, ETS Research Report in press.

Freedle, R., Kostin, I. , & Schwartz (1987). A comparison of Black/White strategies in solving SAT verbal analogies using a thinking aloud method and a matched ability design, ETS Research Report in press.

Goldman, R. E. and Richards, R. (1974). The SAT prediction of grades for Mexican American versus Anglo-American students of the University of California, Riverside. Journal of Educational Measurement. 11 (2), 129-135.

Goldman, R. E., and Hewitt, B. (1975). An investigation of test bias for Mexican-American college students. Journal of Educational Measurement, 187-196.

Goldman, R. E., and Hewitt, B. (1976). Predicting the success of black, Chicano, oriental and white college students. Journal of Educational Measurement, 13 (2), 109-117.

Goldman, R. E. and Widawski, M. (1976). An analysis of types of errors in the selection of minority college students. Journal of Educational Measurement, 13 (3).

Hale, G. A., Stansfield, C. W., & Duran, R. P. (1984) Summaries of studies involving the Test of English as a Foreign Language, 1963- 1982, TOEFL Research Report #16, Princeton, NJ: Educational Testing Service.

Hunter, R.V. & Slaughter, C. D. (1980). ETS test sensitivity review process. Princeton NJ: Educational Testing Service.

Lee, V. E. & Ekstrom, R. B. (1987). Student access to guidance counseling in high school. American Educational Research Journal, 24, 287-310.

Linn, R. L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. K. Wigdor & W. R. Gardner (Eds.), Ability Testing: Uses, consequences, and controversies, 335-388. Washington, DC: National Academy Press.

Loyd, B. H. (March 1982). Analysis of content-related bias for Anglo and Hispanic students. Paper presented at the annual meeting of the American Educational Research Association, New York.

Messick, S. (1987). Validity. ETS Research Report RR-87-40. Princeton, NJ: Educational Testing Service.

Messick, S., and Jungeblut, A. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89, 191-216.

Olmedo, E.L. (1977). Psychological testing and the Chicano. In J. L. Martinez, Jr. (Ed.), Chicano psychology. NY: Academic Press.

Pennock-Roman, M. (1986). New directions for research on Spanish-language tests and test-item bias. Chapter 7, pp. 193-220 in M. A. Olivas (Ed.), Latino College Students, NY: Teachers College Press.

Pennock-Roman, M. (1986). Fairness in the use of tests for selective admissions of Hispanics. Chapter 9, pp. 246-280 in M. A. Olivas (Ed.), Latino College Students, NY: Teachers College Press.

Pennock-Roman, M. Powers, D. & Perez, M. (October 1987) An evaluation of a kit designed to prepare Hispanic High School students to take the PSAT/NMSQT, College Board Report under revision.

Profiles of College Bound Seniors (1987). Princeton, NJ: College Entrance Examination Board.

Rogers, H. J. & Kulick, E. (April, 1986). An investigation of unexpected differences in item performance between Blacks and Whites taking the SAT. In A. P. Schmitt, & N. J. Dorans, (Eds.) (March 1987), Differential item functioning on the Scholastic Aptitude Test, (RM-87-1), Princeton, NJ: Educational Testing Service.

Sanchez, G.I. (1932a). Group differences and Spanish-speaking children: A critical review. Journal of Applied Psychology, 16, 549-558.

Sanchez, G.I. (1932b). Scores of Spanish-speaking children on repeated tests. Journal of Genetic Psychology, 40, 223-231.

Sanchez, G.I. (1934a). Bilingualism and metal measures: A word of caution. Journal of Applied Psychology, 18, 765-772.

Sanchez, G.I. (1934b). The implications of a basal vocabulary for the measurement of the abilities of bilingual children. Journal of Social Psychology, 5,395-402.

Scheuneman, J.D. (1982). A posteriori analyses of biased items. In R.A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore, Md.: The Johns Hopkins University Press.

Schmitt, A. P. (April, 1986). Unexpected differential item performance of Hispanic examinees. Paper presented at NCME/AERA conference in San Francisco, in Schmitt, & N. J. Dorans, (Eds.) (March 1987), Differential item functioning on the Scholastic Aptitude Test, (RM-87-1), Princeton, NJ: Educational Testing Service. (A later edition of this paper is soon to be published in the Journal of Educational Measurement, 1988.)

Schmitt, A. P. & Bleistein, C. A. (1987). Factors affecting differential item functioning for Black examinees on Scholastic Aptitude Test Analogy Items, (RR-87-23). Princeton, NJ: Educational Testing Service.

Schmitt, A. P. & Dorans, N. J. (August 1987). Differential Item Functioning for Minorit Examinees on the SAT. Paper presented at the Annual Meeting of the American Psychological Association, New York.

Shepard, L. A. (April 1986). Discussant comments on the NCME Symposium, in Schmitt, & N. J. Dorans, (Eds.) (March 1987), Differential item functioning on the Scholastic Aptitude Test, (RM-87-1), Princeton, NJ: Educational Testing Service.

Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.) The handbook of methods for detecting test bias. Baltimore, MD: Johns Hopkins University Press.

Shepard, L. A., Camilli, G. & Williams, D. M. Accounting for statistical artifacts in item bias research. Journal of Educational Statistics, 9, 93-128.

Strenta, A. C. & Elliot, R. (1987). Differential grading standards revisited. Journal of Educational Measurement 24, 281-291.

Warren, J., (1976). Prediction of College Achievement among Mexican-American students in California. College Board Research and Development Report. Princeton, NJ: ETS.

Wright, D. (1986). An empirical comparison of the Mantel-Haenszel and standardization methods of detecting differential item performance. Paper presented at the annual meeting of the National Council on Measurement in Education in San Franciso, April, 1986.

Zieky, M. (September 1987). Procedures for use of differential item difficulty statistics in test development. (Memorandum for ETS test developers.)

Footnotes

[1] In operational item analyses, the cutoff value used to flag items for review is expressed in units of the Mantel-Haenszel delta difference index (abbreviated MH D-Diff). By consensus of psychometricians at ETS and outside consultants, the cutoff value has been set equal to or higher than 1.5, provided that the index is also statistically significantly different from 1.00 (Zieky, September, 1987). Since the standardization index is used primarily in research and not in operational item analyses, it has not been necessary to derive the cutoff value for the standardization indexes that is equivalent to the one for MH D-Diff. A general solution to the functional relationship between the standardization index and the Mantel-Haenszel has not been worked out. However, the cutoff for the standardization index that would be equivalent to the 1.5 Mantel-Haenszel cutoff can be estimated through the results of an empirical study by David Wright (1986). This study gives correlations and descriptive statistics that allow us to estimate roughly the regression of the indexes from the standardization method on the Mantel-Haenszel index, and vice-versa, although it is not clear that this relationship generalizes to samples other than one used by Wright. Using this rough approximation, I found that a cutoff of 1.5 in the MH D-Diff would be approximately equal to a D_{STD} of .11. Thus, the cutoff value that Schmitt used was slightly less half the size of the estimate for the usual cutoff for operational analyses. Her cutoff would be approximately equal to a MH D-Diff of .68, which flags more items as potentially discrepant than the cutoff of 1.5.

[2] The kit is expected to be available by the middle of the summer of 1988. Copies can be ordered by writing to: College Board Publication Services, 45 Columbus Ave., New York, 10023-6992.

TABLE 1

SAT® Averages by Ethnic Group, 1976–1985, 1987

SAT Verbal

	1976*	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
American Indian	388	390	387	386	390	391	388	388	390	392	NA	393
Asian American	414	405	401	396	396	397	398	395	398	404	NA	405
Black	332	330	332	330	330	332	341	339	342	346	NA	351
Mexican American	371	370	370	370	372	373	377	375	376	382	NA	379
Puerto Rican	364	355	349	345	350	353	360	358	358	368	NA	360
Other Hispanic	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	387
White	451	448	446	444	442	442	444	443	445	449	NA	447
Other	410	402	399	393	394	388	392	386	388	391	NA	405
All Students	431	429	429	427	424	424	426	425	426	431	431	430
All Men	433	431	433	431	428	430	431	430	433	437	437	435
All Women	430	427	425	423	420	418	421	420	420	425	426	425

SAT Mathematical

	1976*	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
American Indian	420	421	419	421	426	425	424	425	427	428	NA	432
Asian American	518	514	510	511	509	513	513	514	519	518	NA	521
Black	354	357	354	358	360	362	366	369	373	376	NA	377
Mexican American	410	408	402	410	413	415	416	417	420	426	NA	424
Puerto Rican	401	397	388	388	394	398	403	403	405	409	NA	400
Other Hispanic	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	432
White	493	489	485	483	482	483	483	484	487	490	NA	489
Other	458	457	450	447	449	447	449	446	450	448	NA	455
All Students	472	470	468	467	466	466	467	468	471	475	475	476
All Men	497	497	494	493	491	492	493	493	495	499	501	500
All Women	446	445	444	443	443	443	443	445	449	452	451	453

*1976 is the first year for which SAT scores by ethnic group are available. They were not available for 1986 due to changes in the Student Descriptive Questionnaire (SDQ), that students complete when they register for the tests. The SDQ question on ethnic background was changed to include the "Other Hispanic" category for 1987.

Table 2

SAT Means and Percent Distribution by High School GPA and Number of Academic Courses

	Mean SAT-V				Mean SAT-M				% Breakdown			
	NH White	LA	MA	PR	NH White	LA	MA	PR	NH White	LA	MA	PR
High School GPA												
(1) A+	566	525	501	421	632	587	579	510	4%	3%	3%	3%
(2) A	523	468	443	408	581	328	506	474	11%	9%	9%	10%
(3) A-	493	439	422	406	547	497	479	459	13%	11%	12%	10%
(4) B	434	380	371	356	473	423	413	393	54%	57%	57%	56%
(5) C	381	335	336	320	408	370	366	348	17%	20%	19%	21%
(6) D or below	370	332	316	306	390	348	348	320	0%	1%	0%	1%
Difference in means between (1) and (5)	185	190	165	101	224	217	213	162				
Total years of Study in Six Academic Areas												
(a) 20 or more years	493	434	435	402	539	487	482	453	35%	32%	16%	32%
(b) 19 or 19 1/2 years	463	414	411	381	510	460	458	426	13%	13%	10%	12%
(c) 18 or 18 1/2 years	447	399	401	366	492	443	449	404	13%	14%	13%	12%
(d) 17 or 17 1/2 years	430	377	392	358	472	419	439	394	11%	13%	15%	11%
(e) 16 or 16 1/2 years	414	363	369	342	453	402	413	372	9%	10%	14%	10%
(f) 15 or 15 1/2 years	401	348	358	328	437	388	401	360	7%	7%	11%	8%
(g) Fewer than 15 years	378	318	331	309	411	364	371	340	12%	13%	20%	15%
Difference in means between (a) and (g)	115	116	104	93	128	123	111	113				
Overall Mean for Group	447	387	379	360	489	432	424	400				
Overall Number with SAT Scores	788,613	18,895	20,714	10,304	788,613	18,895	20,714	10,304				

Note: From National College Bound Seniors: 1987 SAT Profile, The College Board.

NH White=Non-Hispanic White; LA=Latin American other than MA, PR; MA=Mexican American; PR=Puerto Rican.

BEST COPY AVAILABLE

Table 3

SAT Means and Percent Distribution by Type and Number of High School Mathematics Courses

	Mean SAT-V				Mean SAT-M				% Breakdown			
	NH White	LA	MA	PR	NH White	LA	MA	PR	NH White	LA	MA	PR
Type of Math Course												
(1) Algebra	447	388	380	361	490	434	424	402	97%	97%	98%	96%
(2) Geometry	453	395	388	368	498	443	435	412	94%	91%	90%	88%
(3) Trigonometry	475	418	414	388	535	479	480	445	57%	55%	44%	51%
(4) Precalculus	501	446	438	411	573	520	512	485	28%	25%	19%	24%
(5) Calculus	522	463	448	446	608	549	541	534	19%	15%	12%	10%
(6) Computer Math	465	409	397	386	523	470	454	435	26%	22%	19%	21%
(7) Other Math Courses	423	369	362	343	458	411	402	375	24%	26%	24%	26%
Difference in means:												
between (3) and (1)	28	30	34	27	45	45	56	43				
(4) and (1)	54	58	58	50	83	86	88	83				
Years of Study in Mathematics												
(a) More than 4 years	487	418	417	393	568	502	502	465	14%	14%	11%	11%
(b) 4 years	457	403	393	374	508	454	447	424	50%	47%	43%	45%
(c) 3 or 3 1/2 years	424	368	367	345	444	396	396	373	29%	31%	35%	32%
(d) 2 or 2 1/2 years	394	331	337	324	394	346	350	333	7%	8%	11%	11%
(e) 1 to 1 1/2 years	363	297	308	300	344	316	323	313	0%	1%	1%	1%
(f) Less than 1 year	389	313	301	315	413	337	360	347	0%	0%	0%	0%
Difference in means between (a) and (d)	93	87	80	69	174	156	179	132				

Note: From National College Bound Seniors: 1987 SAT Profile, The College Board.

NH White=Non-Hispanic White; LA=Latin American other than MA, PR; MA=Mexican American; PR=Puerto Rican.

Table 4

SAT Means and Group Differences Broken Down by Language Background and Parental Education

	SAT-V	Difference in SAT-V Mean Compared to White, NH			SAT-M	Difference in SAT-M Mean Compared to White, NH			Percentage Breakdown			
	NH White	LA	MA	PR	NH White	LA	MA	PR	NH White	LA	MA	PR
Total Subpopulation	447	-60	-68	-87	489	-57	-65	-89	100%	100%	100%	100%
Language First Learned												
English First	449	-28	-46	-48	491	-34	-48	-65	94%	21%	44%	23%
English and Another	420	-25	-55	-69	460	-30	-53	-80	5%	41%	40%	45%
Other Language	396	-36	-46	-53	492	-70	-80	-83	1%	38%	16%	32%
Highest Level of Parental Education												
(1) No H.S. Diploma	385	-53	-42	-61	425	-45	-29	-70	2%	19%	27%	18%
(2) H.S. Diploma	419	-41	-60	-62	459	-44	-39	-71	37%	35%	43%	40%
(3) Associate Degree	428	-35	-36	-74	470	-31	-34	-71	7%	6%	7%	7%
(4) Bachelor's Degree	459	-45	-45	-89	503	-41	-49	-78	29%	18%	13%	17%
(5) Graduate Degree	486	-56	-54	-88	529	-47	-59	-77	25%	22%	10%	17%
Difference in means between (5) and (1)	101	98	89	74	104	102	74	97				

Note: From National College Bound Seniors: 1987 SAT Profile, The College Board.

NH White=Non-Hispanic White; LA=Latin American other than MA, PR; MA=Mexican American; PR=Puerto Rican.

BEST COPY AVAILABLE